

# B-Tagging at CDF and DØ Lessons for LHC

Thomas Wright<sup>1</sup>

*University of Michigan  
Ann Arbor, MI 48109*

## Abstract.

The identification of jets resulting from the fragmentation and hadronization of  $b$  quarks is an important part of high- $p_T$  collider physics. The methods used by the CDF and DØ collaborations to perform this identification are described, including the calibration of the efficiencies and fake rates. Some thoughts on the application of these methods in the LHC environment are also presented.

**Keywords:** b-tagging, displaced vertex, jet probability, soft lepton

**PACS:** 13.20.He, 13.87-a, 13.87.Fh

## INTRODUCTION

The identification or “ $b$ -tagging” of jets resulting from the fragmentation and hadronization of  $b$  quarks is an important part of high- $p_T$  collider physics. Some examples include the study of top quark production or searches for low-mass Higgs bosons in the dominant  $b\bar{b}$  decay mode.

Jets containing  $b$  hadrons have distinctive properties that are useful in distinguishing them from other types of jets. One is a relatively high rate of lepton production from semileptonic decays. Because the fragmentation is hard and the  $b$  hadrons retain about 70% of the original  $b$  quark momentum, the leptons will generally have high  $p_T$  relative to the jet  $p_T$ , which makes them easier to identify and separate from lepton sources in generic jets such as decays in flight of  $\pi$ 's or  $K$ 's. The large mass of  $b$  hadrons also helps, as leptons from  $B$  decays will have  $\sim 1$  GeV/ $c$  of  $p_T$  relative to the jet axis, while leptons and fakes in generic jets tend to be more closely aligned with the jet.

A second property of  $b$  hadrons useful for tagging is their long lifetime, with  $c\tau \sim 450$   $\mu\text{m}$ . A  $b$  hadron with  $p_T = 50$  GeV/ $c$  will then fly on average almost half a centimeter before decaying. The large mass of the hadron produces enough opening angle that the daughter particles can have sizable impact parameters with respect to the  $b$  hadron point of origin. These particles also have high  $p_T$ , which reduces the effects of multiple scattering and allows these impact parameters to be measured with good resolution.

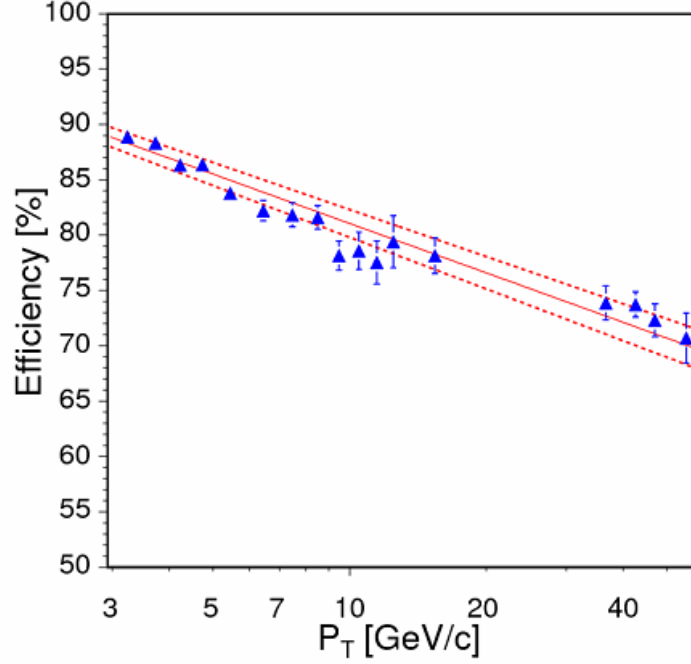
The CDF and DØ detectors are described in Refs. [1, 2]. Both feature a high-efficiency central tracker with good momentum resolution, surrounding a silicon strip detector for precise position measurements. Impact parameter resolutions are typically 40-60  $\mu\text{m}$  depending on the track  $p_T$ , including a 30  $\mu\text{m}$  contribution from the beam width.

## SOFT LEPTON TAGGING

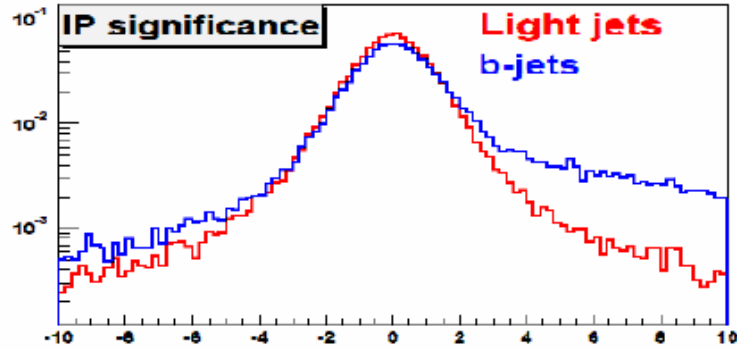
As previously mentioned, the presence of leptons is a good signature of the presence of  $b$  hadrons in a jet. The key is to define an identification algorithm that maintains good performance even in the busy environment around the center of the jet. That means that quantities which are typically used for high- $p_T$  lepton selection, such as calorimeter energy deposition consistent with electrons or muons, cannot generally be used because of the presence of other particles nearby. For this reason muons are somewhat preferred, as most of the identification occurs outside of the calorimeter after the surrounding particles have ranged out. Both CDF and DØ have published results using a soft muon tagger [3, 4]. Figure 1 shows the per-muon efficiency of the CDF soft muon tagger as a function of the muon  $p_T$ , measured using the second legs of  $J/\psi$  and  $Z$  decays. When the branching ratios and  $p_T$  spectra are included the net tagging

---

<sup>1</sup> Speaker, on behalf of the CDF and DØ collaborations



**FIGURE 1.** Efficiency of the CDF soft muon identification as measured from  $J/\psi$  and  $Z$  decays, as a function of the muon  $p_T$ .

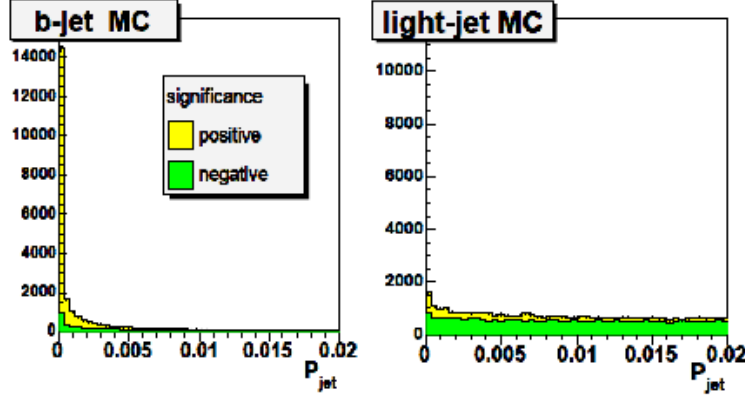


**FIGURE 2.** Impact parameter significance in light-flavor and  $b$  jets at DØ.

efficiency per  $b$  jet is in the 10% range. Fake rates measured from generic jet samples are about 0.5% per muon candidate.

## LIFETIME TAGGING

Tagging algorithms which exploit the long lifetime of  $b$  hadrons have an advantage over the lepton taggers in that they are more inclusive and not limited by semileptonic branching ratios. The basic ingredient of a lifetime tagger is measuring the impact parameters of the tracks within a jet. Figure 2 shows the distributions of impact parameter significance (impact parameter divided by its estimated uncertainty) for light-flavor and  $b$  jets at DØ. The impact parameters are signed such that tracks which cross the jet axis behind the primary vertex relative to the jet direction are negative.  $b$ -jets show a clear excess of tracks with significant positive displacement.



**FIGURE 3.** Probability distributions for  $b$  jets (left) and light jets (right), using tracks with positive or negative impact parameters.

One way to use the impact parameter information is to form the joint probability for all tracks in the jet to have originated from the event primary vertex. This is done using distributions similar to the “light jets” curve in Figure 2 as probability density functions. The tracks with high impact parameter significance which occur in  $b$ -jets will cause this joint probability to peak at low values, as shown in Figure 3. The probability cut can be tuned to obtain a desired purity or efficiency. More information on the CDF and DØ implementations of this algorithm can be found in Refs. [5, 6].

Another way to use the high impact parameter tracks is to require that they be consistent with originating from a secondary displaced vertex. Typically, tracks above a  $p_T$  threshold (1-2 GeV/ $c$ ) and with high impact parameter significance (2-3) are fit to a common vertex. A pruning algorithm removes tracks with high  $\chi^2$  contribution until a set threshold is reached or there are no tracks left. Finally, a cut on the vertex displacement significance is applied to obtain the desired fake rate. Both experiments have implementations of this algorithm [7, 8, 9].

## Efficiency Measurement

Because of the difficulty of modeling the impact parameter distributions in a detector simulation, it is important to measure the efficiency of the taggers from real data. Both experiments use inclusive muon-triggered samples for this purpose. The muon not only enhances the  $b$ -fraction of the data, but its distinctive  $p_T$  relative to the jet axis in  $b$  jets allows that  $b$ -fraction to be measured. Figure 4 shows distributions of this relative  $p_T$  for untagged and tagged jets containing a muon in the CDF data. By fitting templates for the  $b$  and non- $b$  components, the numbers of untagged and tagged  $b$  jets and hence the  $b$ -tagging efficiency can be found. DØ use a similar method, although instead of fitting the  $p_T$  distribution it is split into two bins and the efficiency solved for algebraically. CDF have a second method using an electron sample, with the non- $b$  component inferred from the rate of identified conversion pairs.

Because these jets containing leptons are not representative of generic  $b$ -jets, the efficiencies measured in these samples cannot be used directly. Instead, samples of simulated events passing the same cuts are generated, and a ratio of  $b$ -tagging efficiencies between data and simulation, or “scale factor” is derived. This scale factor can then be used to correct the  $b$ -tagging efficiency in any simulated sample to match the data. In practice the scale factor is a function of jet  $E_T$  and  $\eta$ . Figure 5 shows some typical corrected efficiency parametrizations for the DØ jet probability tagger.

## Fake Rate Measurement

As with the efficiency, the tagger fake rates need to be determined from the data. Generally this is done using tracks with negative impact parameters, which are unaffected by the presence of heavy flavor. The joint probability can be computed using negative tracks instead of positive ones as shown in Figure 3, or for the displaced vertex taggers the rate at which vertices are found behind the primary vertex with respect to the jet direction can be used as an estimate of the fake rate.

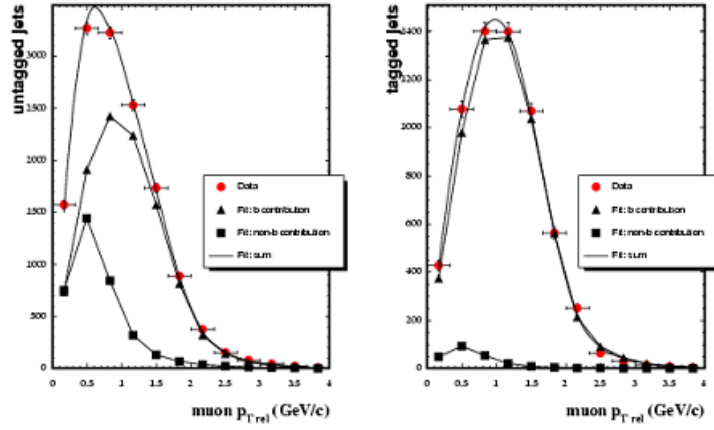


FIGURE 4. Distributions of muon  $p_T$  relative to the jet axis at CDF, for untagged (left) and tagged (right) jets.

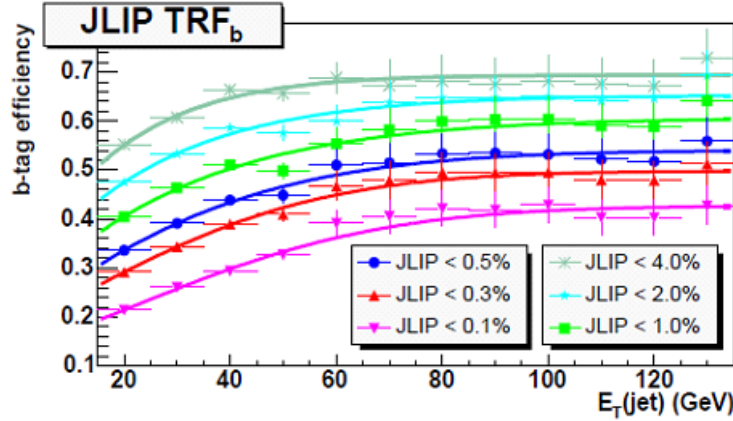
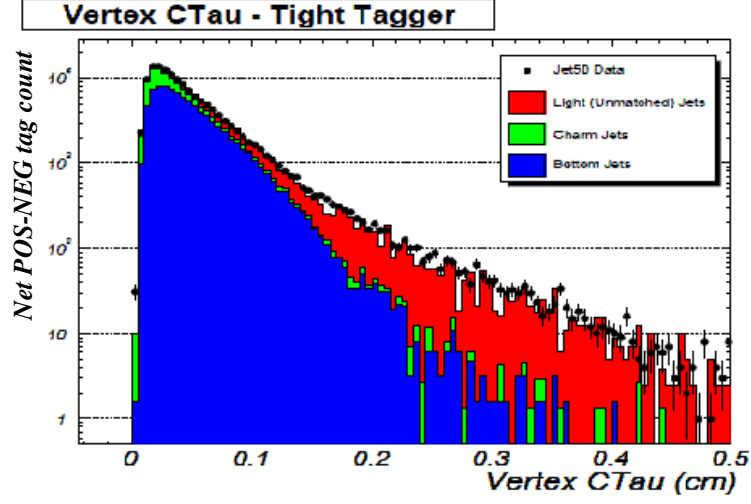


FIGURE 5. Parametrizations of the  $b$ -tagging efficiency for the  $D\bar{D}$  jet probability tagger, derived from simulation and corrected to match the data.

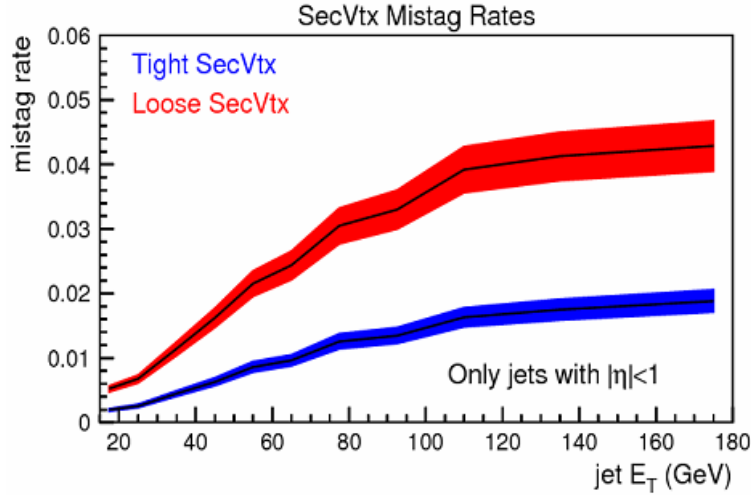
These estimates account for fake tags due to misreconstructed tracks, however they do not include tags from  $K_S/\Lambda$  tracks surviving the removal cuts or from interactions with the detector material, as these will produce preferentially positive tracks and vertices. These effects can be estimated using the pseudo- $c\tau$  (defined as  $L_{xy} \times M_{VTX}/P_{T,VTX}$ ) distribution of vertices, after subtracting the symmetric fake component derived from the negative tags, and fitting for the very long lifetime component as shown in Figure 6 for the CDF displaced vertex tagger. Based on this fit the negative tag rate must be scaled up by 30% to get the positive fake tag rate, shown in Figure 7 as a function of jet  $E_T$ .

## MULTIVARIATE TAGGERS

Rather than simply cutting on the joint probability or the vertex significance, a multivariate discriminant derived from these quantities and other properties of the tags, such as the invariant mass of the tracks in the vertex, the number of tracks in the vertex, etc. can be constructed and used to select jets. Both experiments have released preliminary results [10, 11] using such taggers. Including more information allows for either a more efficient or higher-purity selection than is possible with the single-variable taggers, as shown in Figure 8 for the  $D\bar{D}$  multivariate tagger.



**FIGURE 6.** Net distribution of pseudo- $c\tau$  for the CDF displaced vertex tagger, derived by subtracting the inverted negative tagged distribution from the positive tagged one.



**FIGURE 7.** Postive fake tag rates for two tunes of the CDF displaced vertex tagger.

## SUMMARY

CDF and DØ have developed high-performance and well-understood tools for the identification of  $b$  jets, and have used them to publish many important physics results. These tools will also be important at the Large Hadron Collider when it starts running next year. Much of the experience gained at the Tevatron should be beneficial in commissioning, such as how to choose track quality cuts, the importance of a well-tuned simulation, and triggering strategies for the data samples necessary for characterization of the tagger performance. The large leap in energy and luminosity will certainly present additional challenges, but the physics potential guarantees they will be solved.

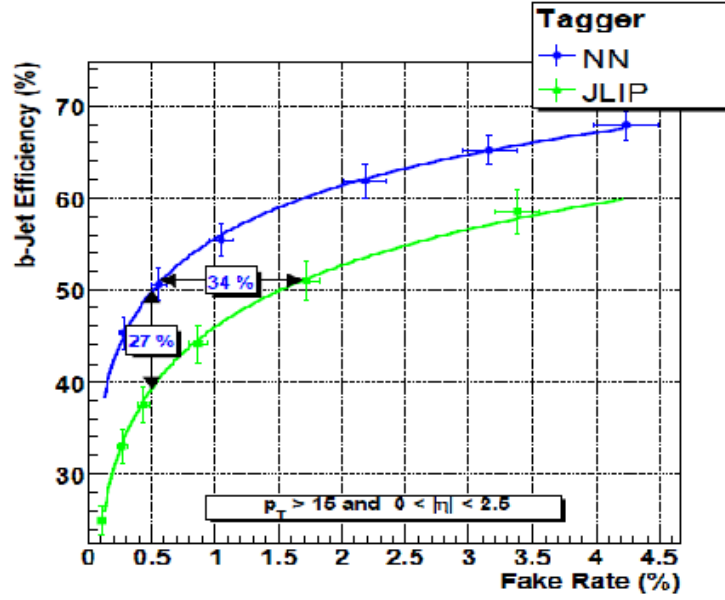


FIGURE 8. Efficiency and purity for the DØ multivariate tagger, compared to their joint probability algorithm.

## REFERENCES

1. D. Acosta *et al.* [CDF Collaboration], “Measurement of the  $J/\psi$  meson and  $b$ -hadron production cross sections in  $p\bar{p}$  collisions at  $\sqrt{s} = 1960$  GeV,” Phys. Rev. D **71**, 032001 (2005) [arXiv:hep-ex/0412071].
2. V. M. Abazov *et al.* [DØ Collaboration], “The upgraded DØ detector,” submitted to Nucl. Instrum. Methods A, [arXiv:physics/0507191].
3. D. Acosta *et al.* [CDF Collaboration], “Measurement of the  $t\bar{t}$  production cross section in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV using lepton plus jets events with semileptonic B decays to muons,” Phys. Rev. D **72**, 032002 (2005) [arXiv:hep-ex/0506001].
4. V. M. Abazov *et al.* [DØ Collaboration], “A search for anomalous heavy-flavor quark production in association with W bosons,” Phys. Rev. Lett. **94**, 152002 (2005) [arXiv:hep-ex/0411084].
5. A. Abulencia *et al.* [CDF Collaboration], “Measurement of the  $t\bar{t}$  production cross section in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$ -TeV using lepton + jets events with jet probability  $b$ -tagging,” submitted to Phys. Rev. D, [arXiv:hep-ex/0607035].
6. S. Greder, “ $b$  quark tagging and cross-section measurement in quark pair production at DØ,” FERMILAB-THESIS-2004-28.
7. D. Acosta *et al.* [CDF Collaboration], “Measurement of the  $t\bar{t}$  production cross section in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV using lepton + jets events with secondary vertex  $b$ -tagging,” Phys. Rev. D **71**, 052003 (2005) [arXiv:hep-ex/0410041].
8. A. Abulencia *et al.* [CDF Collaboration], “Measurement of the  $t\bar{t}$  production cross section in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$ -TeV,” to appear in Phys. Rev. Lett., [arXiv:hep-ex/0606017].
9. V. M. Abazov *et al.* [DØ Collaboration], “Measurement of the  $t\bar{t}$  production cross section in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$ -TeV using lepton + jets events with lifetime  $b$ -tagging,” Phys. Lett. B **626**, 35 (2005) [arXiv:hep-ex/0504058].
10. A. Abulencia *et al.* [CDF Collaboration], “Search for standard model Higgs boson production in association with  $W^\pm$  boson at CDF with  $1 \text{ fb}^{-1}$ ,” CDF/ANAL/EXOTIC/PUBLIC/8390, [http://www-cdf.fnal.gov/cdfnotes/cdf8390\\_WH\\_1fb\\_2.pdf](http://www-cdf.fnal.gov/cdfnotes/cdf8390_WH_1fb_2.pdf).
11. V. M. Abazov *et al.* [DØ Collaboration], “Search for neutral Higgs bosons at high  $\tan\beta$  in multi-jet events,” DØnote 5503-CONF, <http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H23/H23.pdf>.